IA on Marketing

(Customer Analytics)

Index

Intro

A good understanding of customers is extremely important for running a successful business. Know your customer is what actually makes all the difference for many companies. This helps them do their best in creating, communicating and delivering their offerings by tailoring them to their customer's needs.

That makes customer analytics the most important part of both marketing analytics and the marketing function of a company in general.

In fact, customer analytics is a very broad area. It may include a wide range of characteristics of customers and their behavior and numerous different outcomes and performance indicators that the business might be interested in.

This is precisely why I've opted to focus on one of the fundamental marketing frameworks known as the STP framework, which stands for segmentation, targeting, and positioning.

Segmentation

Segmentation is the process of dividing a population of potential or existing customers into groups that share similar characteristics.

For example, not everyone likes the same brand of soft drinks. Moreover, not everyone can afford the same brand of soda.

However, based on certain characteristics like gender, age, purchasing power, we could divide our customers into segments where each segment prefers a brand of soda.

Marketing professionals make the case that taste or spending habits are not the only behavioral features that could be generalized for a segment.

Individuals sharing the same segment may exhibit similar reactions to various marketing initiatives. Conversely, those hailing from distinct segments may have diverse responses to these marketing efforts.

Broadly, the attributes utilized for segmentation can be categorized into two primary groups, depending on whether marketers have access to consumer behavior data.

Frequently, during the creation of new products, consumer behavior data is not accessible. Consequently, marketing profesionals primarily depend on demographic and geographic customer information, including factors like age, income, education level, and more.

In alternative scenarios, marketers have the option to incorporate psychographic characteristics into their segmentation approach.

Consider this: certain customers exhibit well-thought-out purchasing behaviors, while others tend to be more spontaneous in their buying decisions.

The psychological category of segmentation holds an advantage. It comes into play when we possess readily available data on customer consumer behavior. This could include historical purchase data, frequency of purchases, timing of purchases, quantities purchased, product ratings, and various other metrics.

Typically, by leveraging these precise criteria, we can categorize customers into more accurate and representative segments.

Targeting

Targeting revolves around assessing the potential profitability of each segment and making choices about which segments to concentrate on.

Marketers might opt to offer their products to every segment, a select few, or just a single segment. They consider factors such as segment size, anticipated growth, and what competitors are offering.

This phase of the framework is also where we determine the various methods to promote our products. For instance, we may choose to target one segment through television advertising and another through online channels.

Regrettably, targeting activities often involve a qualitative examination of consumer perceptions. They delve into psychology and are often constrained by budget considerations.

Positioning

In essence, positioning in marketing revolves around addressing two key questions:

1. What specific product characteristics are required by customers in a particular market segment?

2. How can products be tailored or presented to align closely with the desired characteristics for that segment?

Positioning not only determining the product's features but also strategizing how to present it to customers and selecting the appropriate distribution channels. This entire process is so crucial that it has its own framework known as the marketing mix.

When discussing the marketing mix, we focus on four key elements, often referred to as the Four P's of Marketing: product, price, promotion, and place. Let's take a closer look at each of these components:

1. Product: This encompasses the core attributes of the offering. It includes features, design, branding, and packaging. For instance, the iPhone is known for its sleek design, distinctive branding, and immaculate packaging.

2. Price: Price pertains to the cost of the product and any related pricing decisions. This involves setting the product's price, considering long-term pricing strategies, offering discounts, and establishing payment terms.

3. Promotion: Promotion encompasses all communication and advertising efforts for the product. It goes beyond just discounts and includes activities like TV commercials, flyers, and unique promotions like sending a Tesla into space. Effective promotion involves crafting messages and deciding on the frequency of communication.

    - Sales Promotions: This includes tactics like price reductions, conditional discounts, and display promotions.

    - Feature Promotions: Feature promotions involve distributing and presenting opportunities for product purchase, such as printed ads and product placements in movies.

1. Place: Place refers to the distribution strategy, specifically where and how the product will be offered to consumers. There are three distribution approaches:

    - Intensive Distribution: This approach involves making the product available in many different stores, aiming for widespread accessibility (e.g., Coca-Cola).

    - Selective Distribution: In selective distribution, the product is only available in specific, carefully chosen stores that are deemed most conducive to sales (e.g., iPhones in Apple Stores).

    - Exclusive Distribution: Exclusive distribution means that only one selected brand is sold in a store, often used for luxury items to enhance the brand's image (e.g., Tesla and Rolex stores).

The marketing mix tools are essential for developing a well-rounded marketing strategy that considers product features, pricing, promotion methods, and distribution channels to effectively reach and satisfy the target audience.

Online vs. Physical Stores

There some differences about the data and process applied if it is an online or physical store.

Brand Choice in Online vs. Physical Stores

Online stores provide more data on customer’s interactions with various brands and their prices, making it easier to predict which brand a customer might choose. In physical stores, this information is limited, unless surveillance cameras are used(not viable).

Loyalty Card Usage and Store Visits

Loyalty card usage for physical stores may be a reliable indicator of future purchases. In other hand, monitoring online store visits could be a better predictor of future buying behavior.

Predicting Purchase Quantity

The ability to predict how many units a customer will purchase is generally similar for both physical and online stores, with online stores benefiting from more extensive data.

Online Store Data Advantage

Online stores offer valuable data, such as customer ratings and reviews, which can be used for predictive modeling. This data is typically not available in physical stores.

Price Elasticity

Price elasticity measures how changes in price affect purchasing behavior. It's essential for businesses to find the optimal price point that maximizes revenue.

Mathematical Representation of Price Elasticity

Price elasticity is the percentage change in an economic outcome (e.g., units sold) in response to a 1% change in price. It is calculated by dividing the percentage change in the outcome by the percentage change in price.

Price Elasticity in Marketing
Price elasticity is crucial for understanding customer behavior. It is used to estimate price elasticity for purchase probability, brand choice probability, and purchase quantity. These elasticities help marketers make informed decisions about pricing strategies.

Purchase Probability

Calculating price elasticity of purchase probability helps determine how changes in the aggregate price of a product category affect the likelihood of a customer making a purchase.

Brand Choice Probability

Price elasticity of brand choice probability focuses on how changes in the price of a specific brand impact the likelihood of customers choosing that brand over others.

Purchase Quantity

Price elasticity of purchase quantity examines how variations in product price influence the quantity customers are willing to buy.

These concepts are essential for understanding consumer behavior, optimizing pricing strategies, and making informed marketing decisions.

About the datasets

I work on a digital marketing agency.
The data I have use throughout this TFM is from a client of our marketing agency, it’s comes from a fast moving consumer goods. The data is anonymized by default.

The datasets are based on B2C (Business 2 Client), this means that the clients of our business are individual people rather than B2B (other companies), this is better from datascience point of view because we have more data points.

An example of a fast moving consumer goods business is a supermarket.

People visit supermarkets every day and most types of goods in store are purchased daily too. Therefore, we have lots and lots of data available.

Customer Analytics Segmentation

data_customers.csv

The information contained comes from the purchasing behavior of these 2000 individuals when entering a physical store. All data has been collected through the fidelity cards they use at checkout.

We can see that here we have an ID column and four demographic and geo demographic variables of 2000 individuals.

The first column is an identifier, the individual ID.

The second column is the biological sex of the individual (gender). A value of zero means male and a value of one means female.

The third column is the age in years.

The fourth column annual_purchase is the spend by the customer over last year in euros.

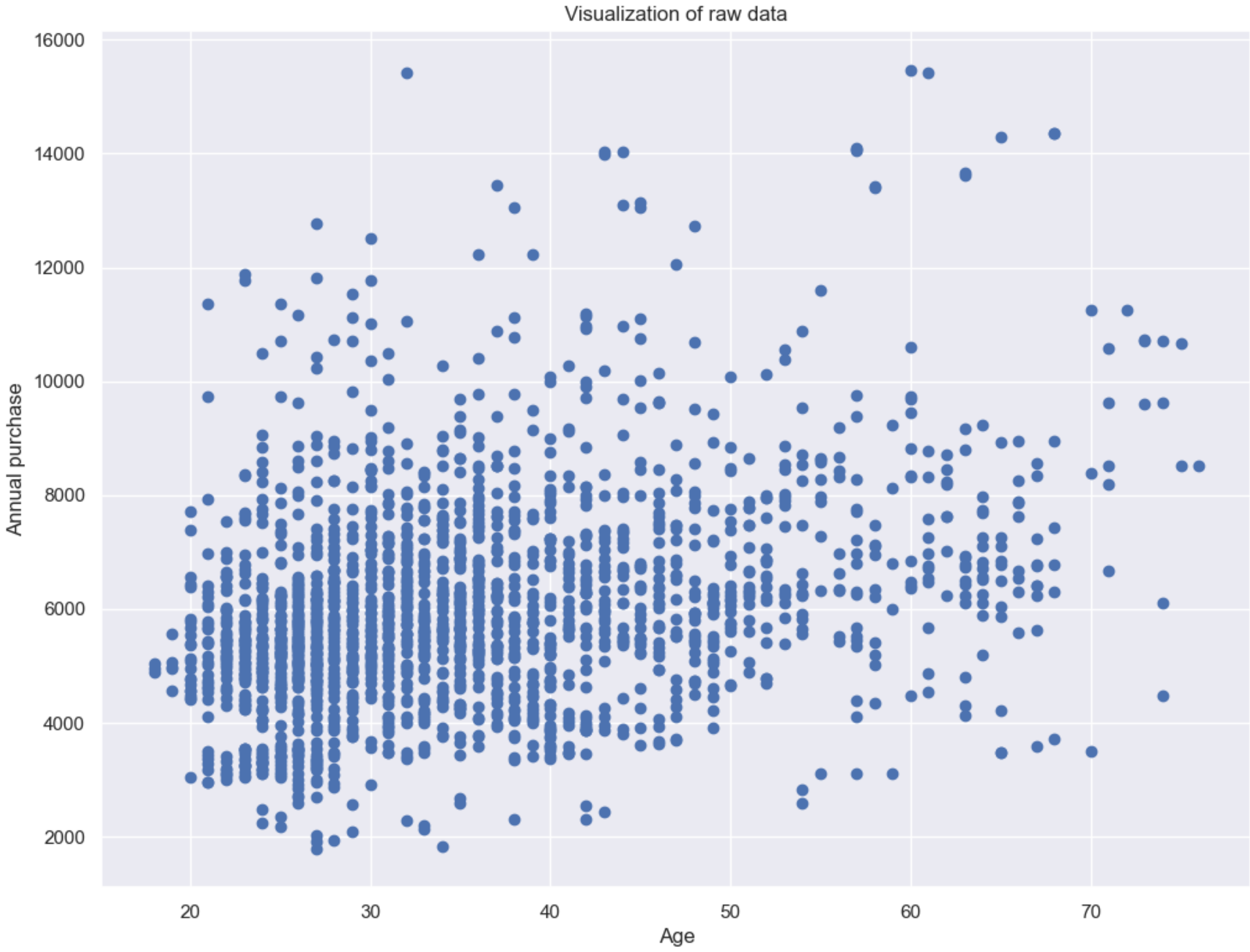
And finally the City_size is the size of the customer's city of residence, a small city that would be encoded with zero, from a mid-sized city with one and a big city two.

	Sex	Age	Annual_purchase	City_size
ID				
100000001	0	67	6234	2
100000002	1	22	7539	2
100000003	0	49	4461	0
100000004	0	45	8579	1
100000005	0	53	7452	1

We can look I there are nulls

```
Sex                0
Age                0
Annual_purchase    0
City_size          0
```

There no nulls on dataset



Data visualization

	Sex	Age	Annual_purchase	City_size
count	2000.000000	2000.000000	2000.000000	2000.000000
mean	0.457000	35.909000	6047.998000	0.739000
std	0.498272	11.719402	1905.433761	0.812533
min	0.000000	18.000000	1792.000000	0.000000
25%	0.000000	27.000000	4883.250000	0.000000
50%	0.000000	33.000000	5778.000000	1.000000
75%	1.000000	42.000000	6904.000000	1.000000
max	1.000000	76.000000	15468.000000	2.000000

In the continuous variables

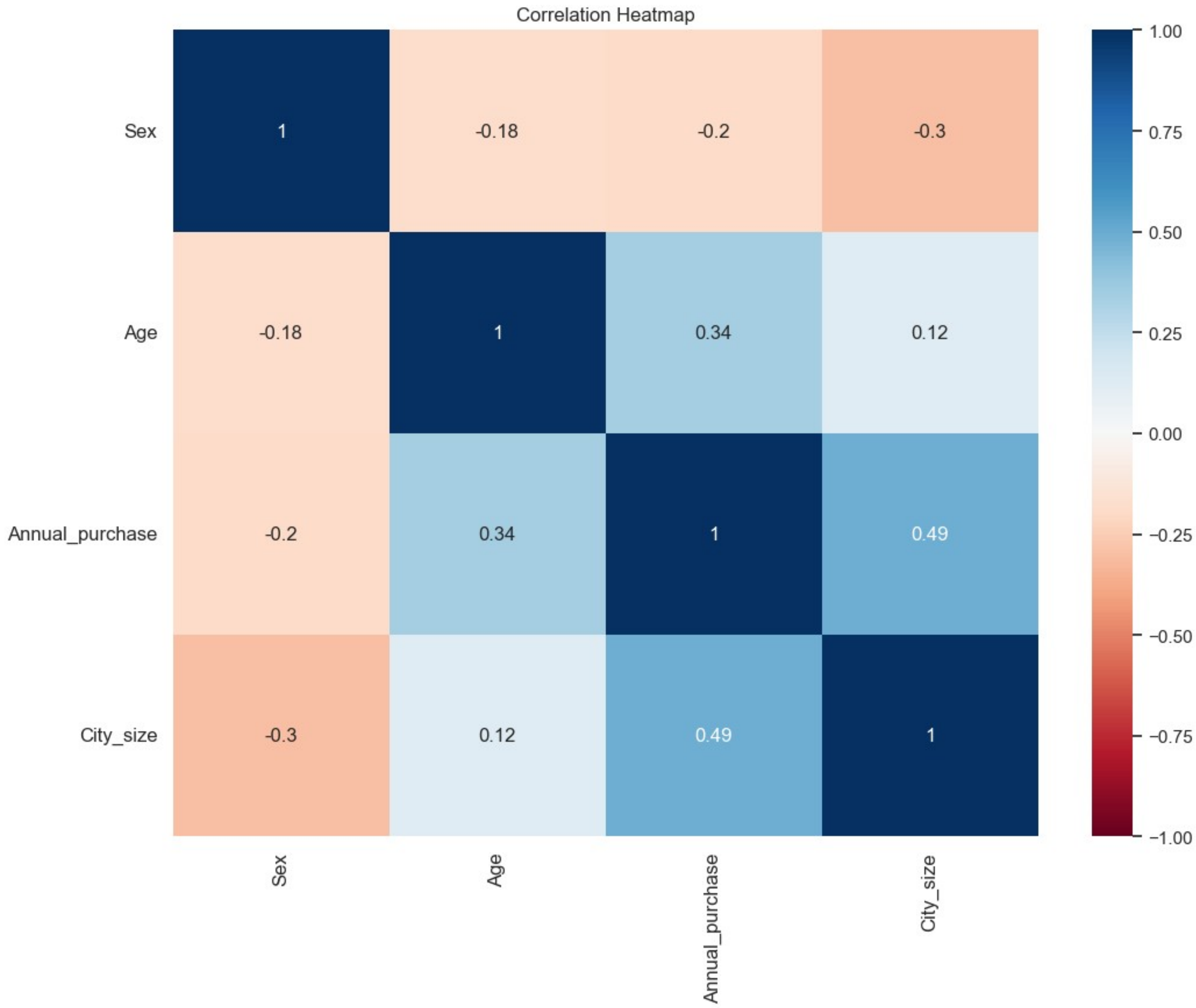
We can see that there are 2000 observations in total.
The average age is about 36 years, while the average annual_purchase is 6047 €
Minimal age of our customers are 18 yo and max 76 yo.
Minimal annual_purchase is 1.792 Euros and max is 15.4468

In the categorical variables

The mean shows nothing more than the proportion of ones.
Let's take biological sex, for instance. Zero means male biological sex and one means female.
So the proportion of women in the data set is 45.7%.

A good way to get an initial understanding of the relationships between the different variables is to explore how they correlate depending on the types of data, you can employ different techniques to quantify this correlation.

For our initial data exploration purposes, we are gonna use Pearson correlation with heatmap.
Pearson is the default approach for most correlation methods in Python.



Generally correlation describes the linear dependency between variables.
It ranges from negative one, one with one indicating very strong positive correlation.
Negative one strong negative.
Collinearity a correlation of zero between two variables means they are not linearly dependent.
The diagonal values show the correlation of a variable with itself. So these values are always one.
The matrix is symmetrical, so the entries over and under the diagonal are mirror images of themselves.
So now the blue or the square between two variables, the higher the positive correlation is between them.
And on the opposite side, the redder the square, the higher the negative correlation is.
With this heat map, we can actually see the strong positive correlation between annual_purchase and city_size and age and annual_purchase.

In simpler terms, the initial step in segmentation involves examining how various characteristics of consumers relate to each other. This process helps in finding commonalities among consumers and grouping them based on these similarities.

As the dataset provided is clean, we don’t need do transformations.

When creating our segmentation models, we focus on comparing and measuring the similarities and differences among individual consumers based on various characteristics that define them. However, we need to be careful when comparing these characteristics.

For instance, consider the age difference between two consumers: one is 20 years old, and the other is 70 years old, resulting in a 50-year difference. While this age gap is substantial, it covers nearly the entire age range in our dataset.

Now, let's also consider their annual purchases. The first consumer spends €10,000 per year, while the second spends only €2,000. This creates an €8,000 difference in their purchase behavior.

If we feed these raw numbers directly into our segmentation models, the algorithms may wrongly emphasize the purchase behavior because the numerical difference of €8,000 greatly outweighs the age difference of 50 years. As a result, the model may not properly consider age as a relevant factor in the segmentation process.

To avoid this issue, we need to standardize or normalize our data, ensuring that all variables are on a comparable scale. This ensures that no single variable dominates the segmentation process due to its numerical magnitude.

Segmentation models.

In essence, cluster analysis originally emerged in anthropology as a means to explain the origins of human beings. Over time, it found applications in various other fields.

The primary objective of clustering is to organize individual observations into groups in such a way that the members within each group are highly similar to each other.

There are two main types of clustering approaches: hierarchical and flat.

Hierarchical Clustering: Hierarchical clustering involves arranging data in a hierarchical or tree-like structure. An example is the taxonomy of the animal kingdom, where the general term "animal" is divided into subgroups like fish, mammals, and birds. Mammals, in turn, can be divided into species that give birth to fully developed offspring, and some species may have subspecies. This hierarchical structure is similar to the organization of clusters.

There are two subtypes of hierarchical clustering:

- Divisive Clustering (Top-Down): This approach starts with all observations in a single cluster and then splits it into smaller clusters. The process continues, potentially until each observation forms its own cluster. Finding the best split requires exploring all possibilities at each step.

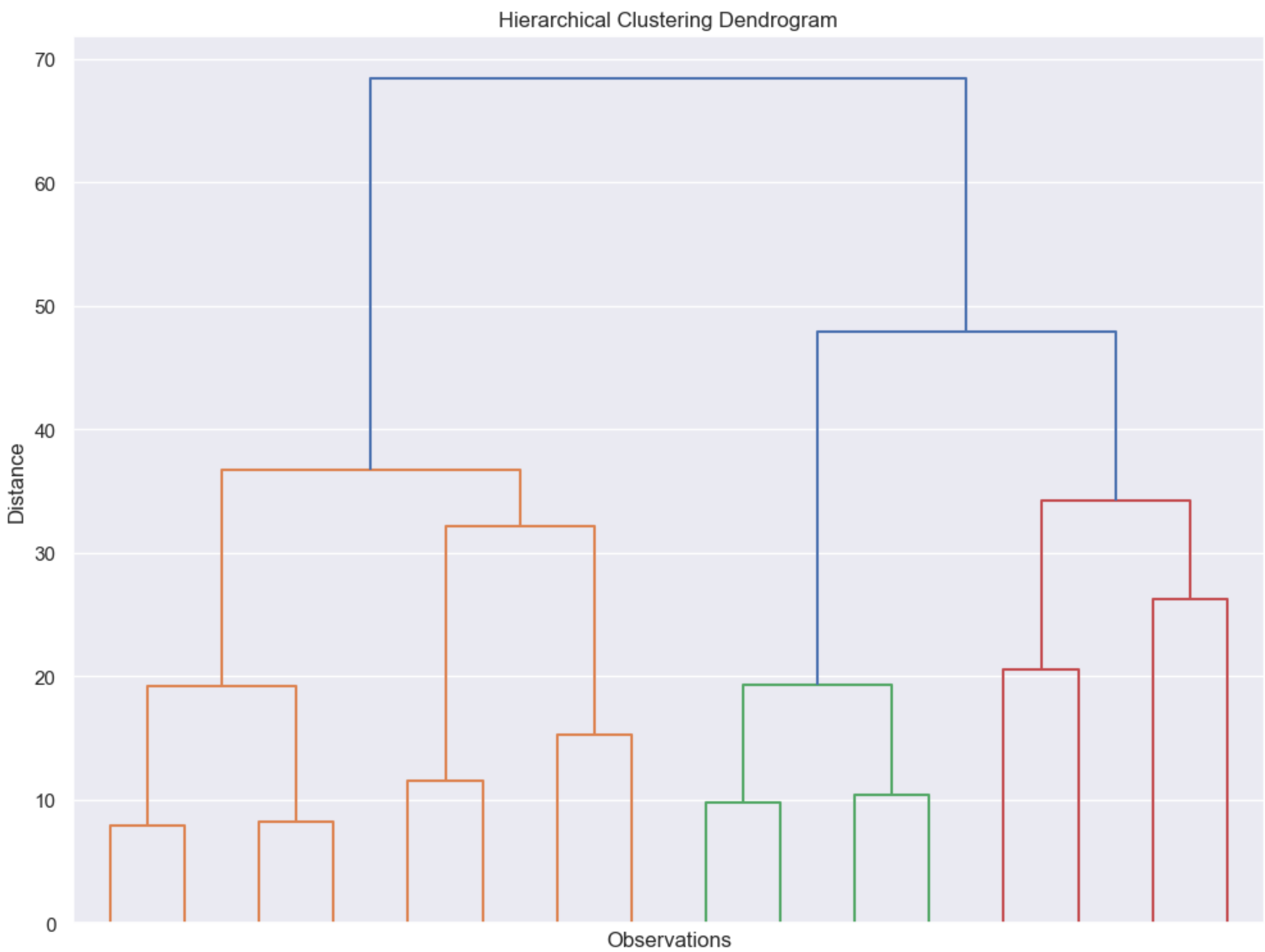
- Agglomerative Clustering (Bottom-Up): Agglomerative clustering begins with individual observations or small clusters, such as different kangaroo and koala species. It then gradually combines or pairs up groups of species until it reaches the highest-level cluster, which, in this example, would be the broader "animal" cluster. Agglomerative and divisive clustering methods should generally yield similar results.

These clustering methods help uncover patterns and relationships within data by grouping similar observations together in a hierarchical manner.

To do the clustering, we calculate the distances between observations
There are different measures we can use, Euclidean distance, the squared, Euclidean distance, Manhattan distance, or maximum distance to just name a few.

However, clustering is about finding the clusters for which points in the same cluster are as similar as possible, while different clusters are as different as possible.
Therefore, we must also consider how to calculate the distances between the clusters as a whole.
We have a lot of options for segmentation problems, such as dividing a population of people.
By far, the most widely used method is the ward method.
The ward method calculates the average of the squares of the distances between clusters, and it is hierarchical clustering and representation.

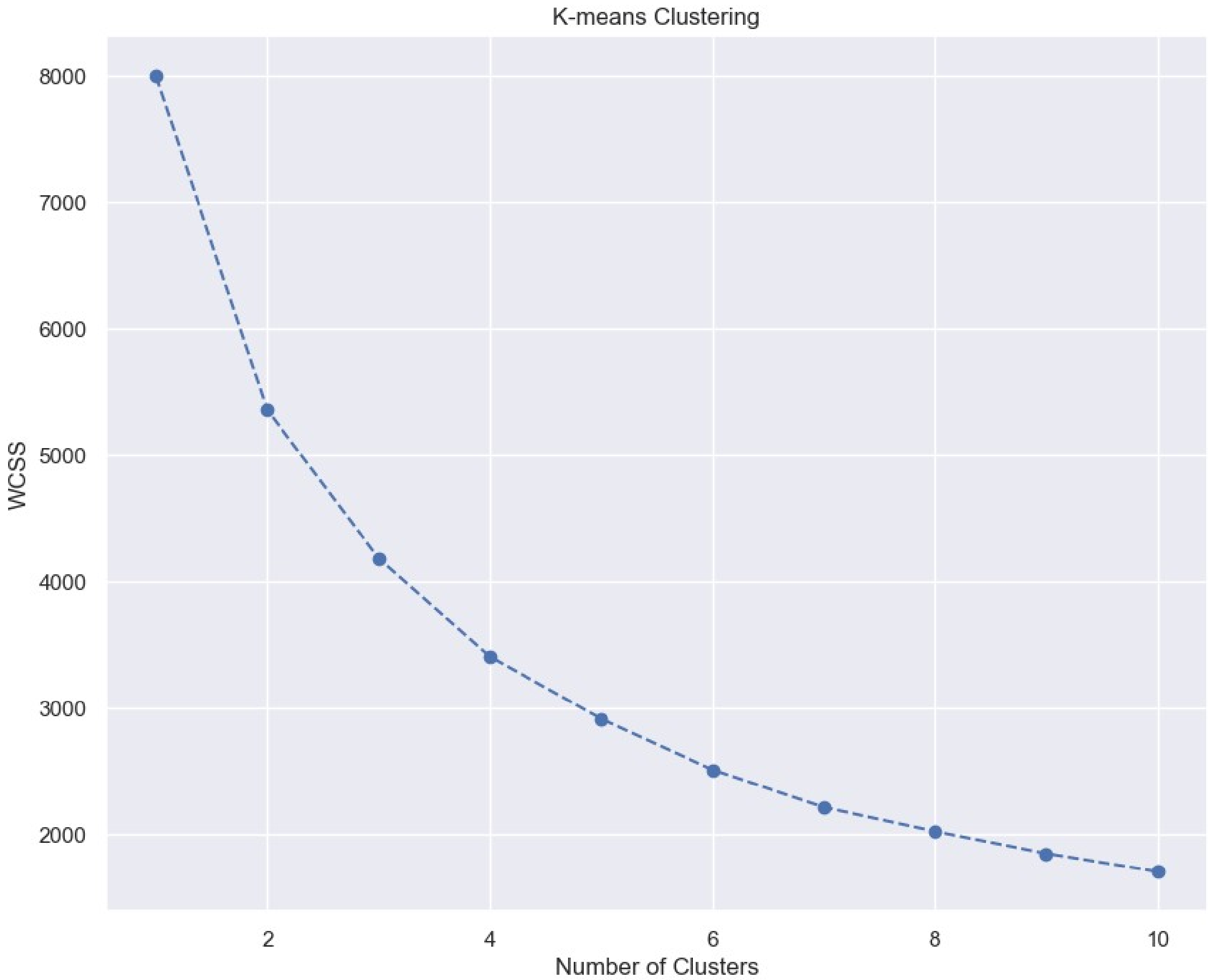
A dendrogram is a tree like hierarchical representation of points.
It provides a nice visual and it's the main tool that is used to assess a hierarchical clustering solution.



We use the Dendrogram for “see” the possible number clusters of our data.
There about 3 or 4 clusters.

Kmeans

There's another method for find the number of clusters, K-means. the K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.



As we see my be 4 clusters on this dataset and apply 4 clusters and group by segment k means and apply the mean for every field, age, sex, etc.

	Sex	Age	Annual_purchase	City_size
Segment K-means				
0	0.302251	55.176849	8202.745981	1.228296
1	1.000000	31.307377	5184.136612	0.213115
2	0.000000	34.478632	5046.702991	0.202991
3	0.179959	31.912065	6929.032720	1.728016

Data analytics

segment 0 have a big annual purchase mean also are oldies than the others and are next to be males.
Segment 1 are females young , lives on small cities the annual purcahse is low
Segment 2 are males young , lives on small cities the annual purcahse is low.
Segment 3 are males young, lives on medium or big cities and spend is medium.

We can do two things the only diference in segment 1 and 2 are the gender, we can join 1 and 2 on one segment or continue with the different segments.

Segment K-means	Sex	Age	Annual_purchase	City_size	N Obs	Prop Obs
0	0.302251	55.176849	8202.745981	1.228296	311	0.1555
1	1.000000	31.307377	5184.136612	0.213115	732	0.3660
2	0.000000	34.478632	5046.702991	0.202991	468	0.2340
3	0.179959	31.912065	6929.032720	1.728016	489	0.2445

We do the count of every segment and, if we join the segment 1 and 2 may unbalance all dataset, we keep separated.

I try and not unbalance the dataset but some as age and annual purchase and city are more homogeneous.

Segment K-means	Sex	Age	Annual_purchase	City_size	N Obs	Prop Obs
0	0.166667	41.558997	7667.613569	1.615044	678	0.3390
1	0.000000	34.721689	5169.262956	0.255278	521	0.2605
2	1.000000	31.898876	5248.649189	0.312110	801	0.4005

I will keep the 4 clusters.

Segment K-means	Sex	Age	Annual_purchase	City_size	N Obs	Prop Obs
mature medium city	0.302251	55.176849	8202.745981	1.228296	311	0.1555
male small city	1.000000	31.307377	5184.136612	0.213115	732	0.3660
female small city	0.000000	34.478632	5046.702991	0.202991	468	0.2340
young big city	0.179959	31.912065	6929.032720	1.728016	489	0.2445

PCA

PCA, or Principal Component Analysis, is a dimensionality reduction technique commonly used in statistics and data analysis. Its primary purpose is to simplify complex data sets while preserving essential information. Here's a breakdown of what PCA is and how it works:

1. Dimensionality Reduction: In many real-world applications, data sets can have a large number of features or variables. PCA aims to reduce the dimensionality of these data sets, meaning it reduces the number of variables or attributes while retaining as much meaningful information as possible.

2. Orthogonal Transformation: PCA achieves dimensionality reduction through an orthogonal transformation. It identifies a set of new variables, called principal components, that are linear combinations of the original features. These principal components are orthogonal to each other, meaning they are uncorrelated.

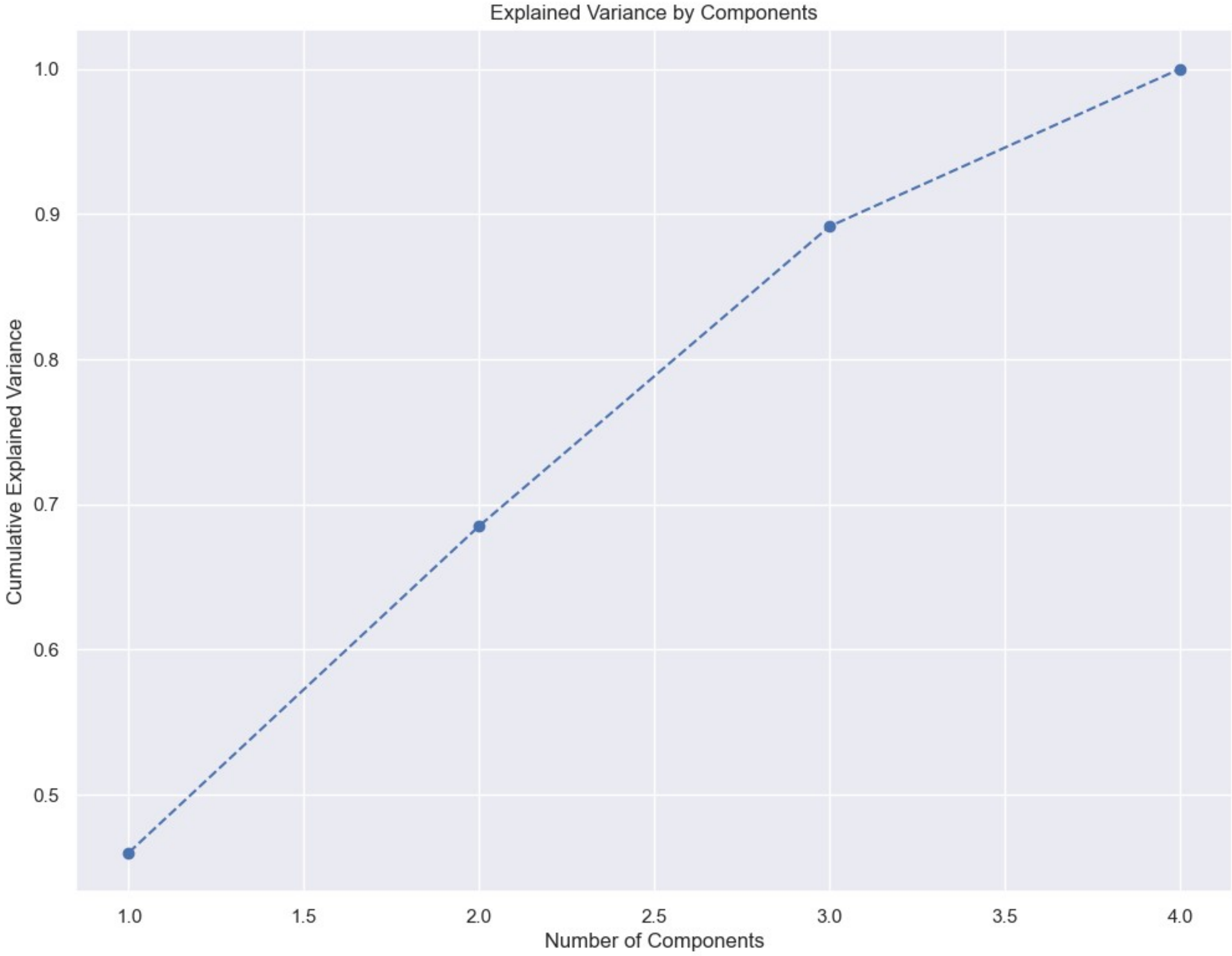
3. Variance Maximization: The principal components are constructed in such a way that the first principal component accounts for the most variance in the data, the second principal component accounts for the second most, and so on. This means that the first few principal components capture the most important information in the data, while the subsequent components capture progressively less important information.

4. Data Compression: PCA allows for data compression because you can represent the data with fewer principal components while retaining most of the original information. This can be useful for tasks like data visualization, where high-dimensional data can be difficult to visualize effectively.

5. Applications: PCA is widely used in various fields, including image compression, data visualization, feature selection, and noise reduction. It is also a fundamental tool in machine learning for preprocessing data before applying algorithms like clustering or classification.

6. Eigenvalues and Eigenvectors: The computation of principal components involves finding the eigenvalues and eigenvectors of the data's covariance matrix. These eigenvalues represent the variance explained by each principal component, and the eigenvectors represent the directions in the original feature space along which the data varies the most.

In summary, PCA is a mathematical technique used to reduce the dimensionality of complex data sets while preserving the most important information. It is a valuable tool for data analysis, visualization, and preprocessing in various fields, including statistics, machine learning, and image processing.
We want to do dimensional reduction to simplify our problem.



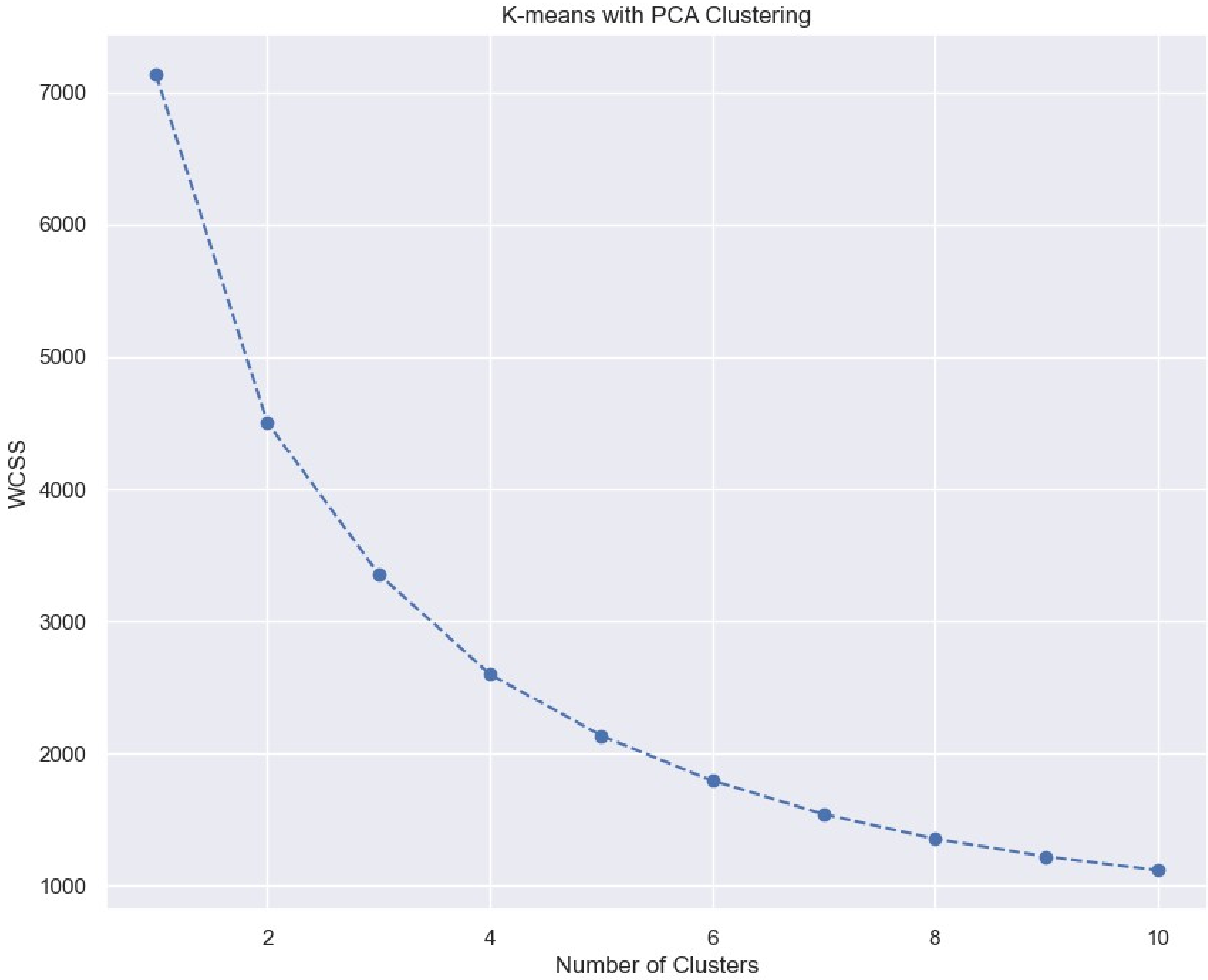
I will select 3 components for better explainability and keep about the 90% of the variance.
We create a headmap for see the correlation with the variables with the components.

We see a heavy

correlation positive between sex and component 3, component 2 and age.
Also we have a less heavy correlation in component 1 and city_size and annual-purchase.
Also we have a less heavy inverse correlation in component1 and sex and component2 and city size.

Component1 explain, its no clear, city size and annual purchase.
Component2 explain age
Component3 explain sex and annual purchase

K-means clustering



We've obtained the K-means clustering solution with four or three clusters.
We are gonna use 3 clusters In this case.

With 4

Segment K-means PCA	Sex	Age	Annual_purchase	City_size	Component 1	Component 2	Component 3
0	0.000000	34.540305	4995.200436	0.204793	-0.348032	-0.178402	-1.169562
1	1.000000	31.261286	5163.177839	0.216142	-1.254219	0.296222	0.561018
2	0.306189	55.371336	8213.361564	1.214984	1.801682	1.106956	0.066458
3	0.176938	32.033797	6972.990060	1.695825	1.040685	-0.943314	0.211376

With 3

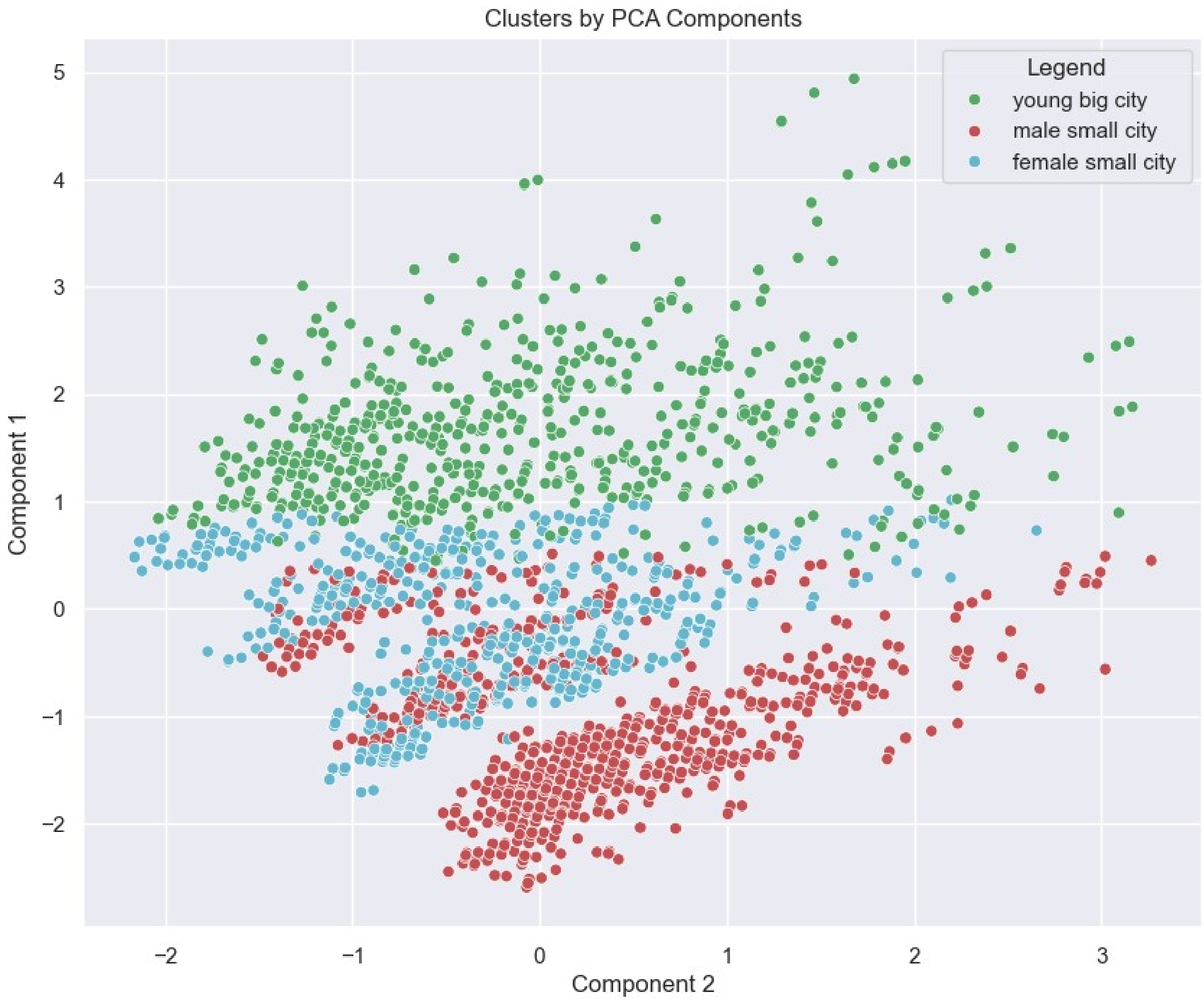
Segment K-means PCA	Sex	Age	Annual_purchase	City_size	Component 1	Component 2	Component 3
0	0.16443	43.474832	7946.484899	1.583893	1.675528	-0.039398	0.149919
1	0.00000	33.952381	5196.200680	0.430272	-0.152834	-0.332802	-1.022653
2	1.00000	31.792892	5275.153186	0.344363	-1.113662	0.268589	0.627412

I think is better 3 because, the figures with the components of PCA we can explain better, and can be less overlap.

There are less clusters we lost mature medium city, but we can with the components what cluster equals.

Component1 on segment0 explain city size and annual purchase, it's young big city.
Component3on segment2 explain male and annual purchase, it's male on little city
Component3 on segment1 explain sex, its female on little city

This is the results of clasificaction on kmeans using PCA



On 3

Young big city is clear
The male small city and female small city more difuminated but not bad.

Purchase Analytics

For purchase analytics we are gonna study another dataset.
This dataset have this columns:

ID	Shows a unique identificator of a customer.
Day	Day when the customer has visited the store
Incidence	Purchase Incidence 0 The customer has not purchased an item from the category of interest 1 The customer has purchased an item from the category of interest
Brand	[1-5] Shows which brand the customer has purchased 0 No brand was purchased
Quantity	Number of items bought by the customer from the product category of interest
Last_Inc_Brand	[1, 5] Shows which brand the customer has purchased on their previous store visit 0 No brand was purchased
Last_Inc_Quantity	Number of items bought by the customer from the product category of interest during their previous store visit
Price_1	Price of an item from Brand 1 on a particular day
Price_2	Price of an item from Brand 2 on a particular day
Price_3	Price of an item from Brand 3 on a particular day
Price_4	Price of an item from Brand 4 on a particular day
Price_5	Price of an item from Brand 5 on a particular day
Promotion_1	Indicator whether Brand 1 was on promotion or not on a particular day 0 There is no promotion 1 There is promotion
Promotion_2	Indicator whether Brand 2 was on promotion or not on a particular day 0 There is no promotion 1 There is promotion
Promotion_3	Indicator whether Brand 3 was on promotion or not on a particular day 0 There is no promotion 1 There is promotion
Promotion_4	Indicator whether Brand 4 was on promotion or not on a particular day 0 There is no promotion 1 There is promotion

Promotion_5	Indicator whether Brand 5 was on promotion or not on a particular day 0 There is no promotion 1 There is promotion
Sex	Biological sex (gender) of a customer. In this dataset there are only 2 different options. 0 male 1 female
Age	The age of the customer in years, calculated as current year minus the year of birth of the customer at the time of creation of the dataset
Annual_purchase	It’s the spend by the customer over last year in euros.
City_size	The size of the city that the customer lives in. 0 small city 1 mid-sized city 2 big city

Each observation is a transaction rather than a customer.
Therefore, several different observations can relate to the same customer.
There are 500 unique individuals and we have information about their purchases on a daily basis.

For instance, you can incentive them with coupon codes or loyalty bonuses and exchange for participation in such an experiment.
Alternatively, you can use scanner panel data similar to the data Nielsen and IRI collect.
Either way, we've got the purchases of 500 customers on our shop.
We know whether the individual had visited the shop and we know all the items they had purchased.

Since this is a lot of information, we will only focus on the purchases of sodas.
All data is related to the purchase of sodas.

Purchase incidents.

This is an indicator variable for whether a product from the category of interest was purchased on the shopping(sodas).

So it is one when the customer visited the shop and purchased one or more sodas and zero when they didn't buy.
Then there are five different brands of sodas available.
So the next column has values from 1 to 5 denoting which brand the customer purchased.
It holds the value of zero if they did not purchase sodas at that time.

Then there is the purchase quantity column, which shows how many chocolate candy bars the customer purchased.

Well, this purchase behavior depends on the efforts of each brand to make itself more attractive for the customers.

In other words, it can be influenced by the marketing mix tools.

The next two columns represent the brand and purchase quantity chose on the previous purchase occasion.

Next is data about the prices of one soda for each of the brands for each day.

We also have an indicator variable about whether a brand was on promotion.

Logically, if the brand was on promotion that day, the indicator variable has a value of one and zero otherwise.

We distinguish between three types of sales promotions or merchandising, price reduction, display and feature.

Our promotion indicator variable reflects any of these promotions.

Okay when collecting the individuals purchases, some geodemographic characteristics were also gathered.

In fact, they are represented by the same four columns we observed in the segmentation part.

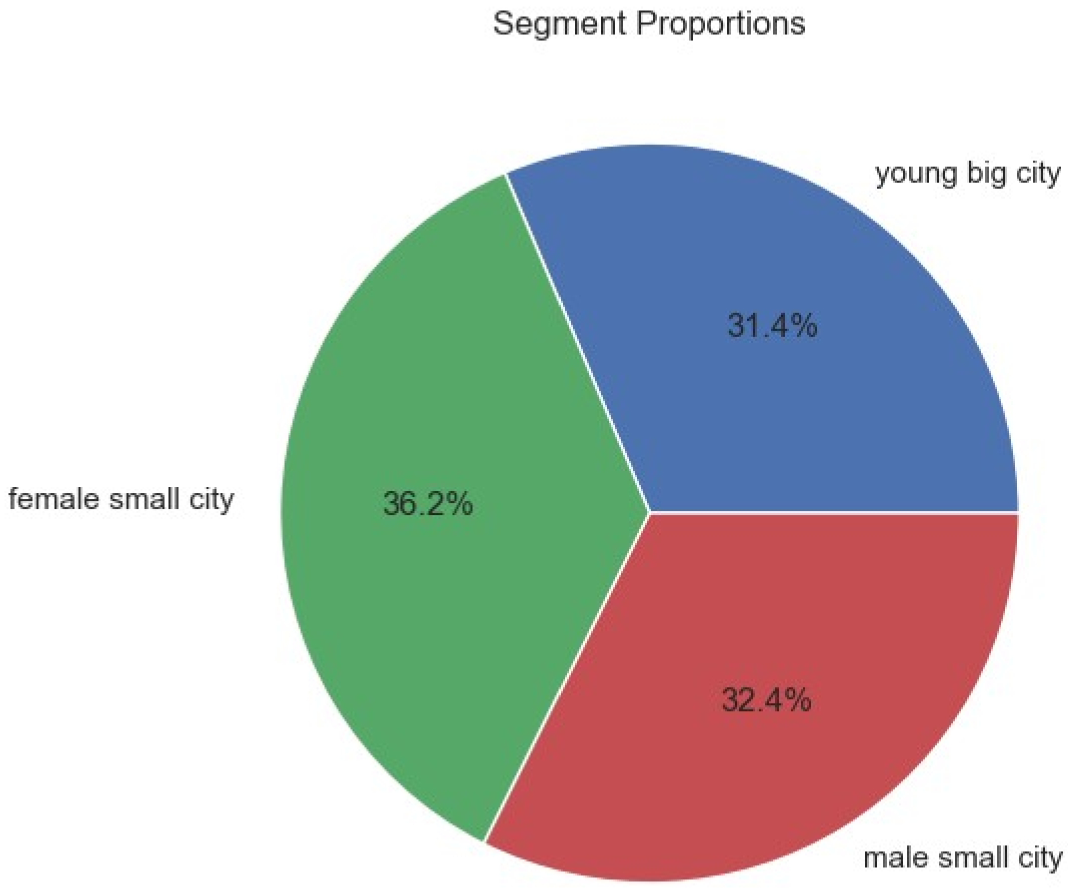
Sex, age, Annual_purchase, City size.

By the way, their set of values are identical to the ones from the segmentation data.

We have clustering new data with the PCA knime clasificaction model.

nc_Quantity	Price_1	Price_2	Price_3	...	Promotion_1	Promotion_2	Promotion_3	Promotion_4	Promotion_5	Sex	Age	Annual_purchase	City_size	Segment
0	1.59	1.87	2.01	...	0	1	0	0	0	0	47	5543	0	1
0	1.51	1.89	1.99	...	0	0	0	0	0	0	47	5543	0	1
0	1.51	1.89	1.99	...	0	0	0	0	0	0	47	5543	0	1
0	1.52	1.89	1.98	...	0	0	0	0	0	0	47	5543	0	1
0	1.52	1.89	1.99	...	0	0	0	0	0	0	47	5543	0	1

We are interested in have a visual idea about how the store visitors are distributed across segments.



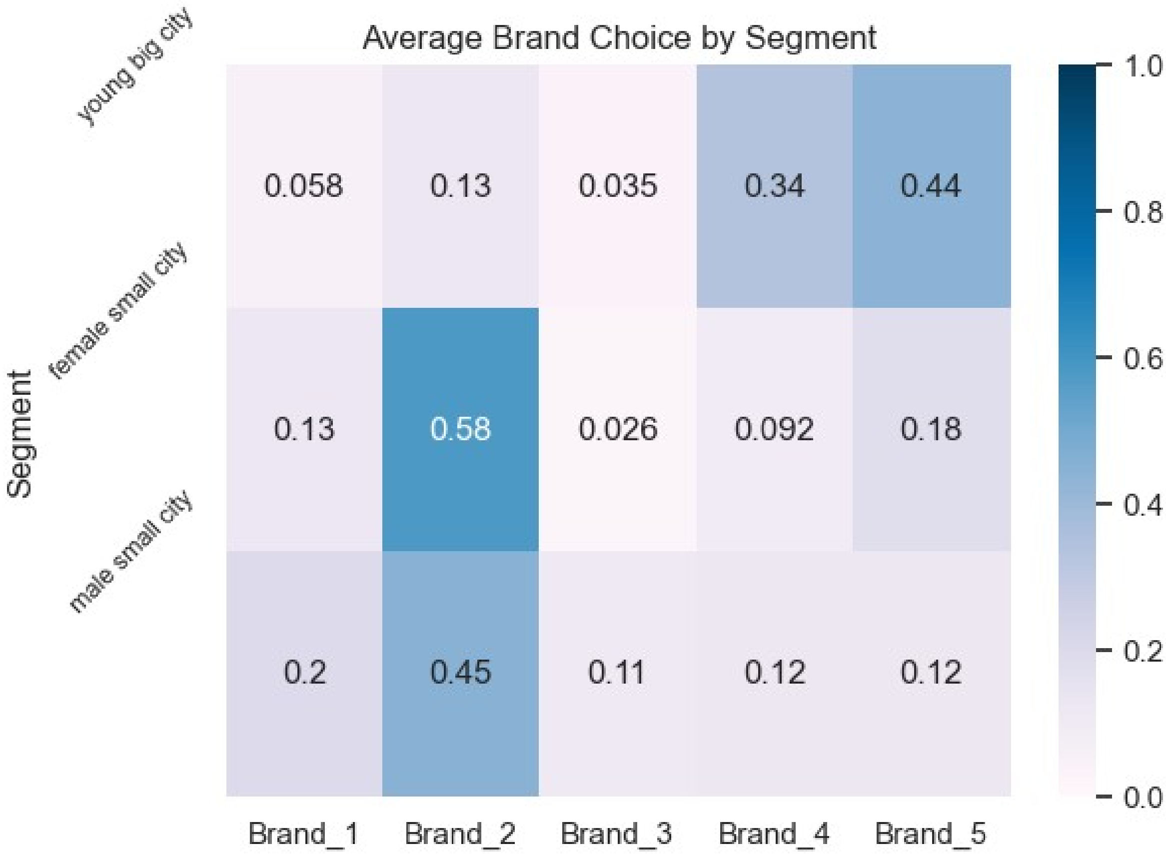
We have the proportions of the segments (number of rows by every segment).
The female small city have the 36% the rest are almost equally distributed arround 32%

We are interested on the average number of visits by segment, for valorate better the purchases.
For example if a segment goes less and buy less, may be necessary do an adjust on the data for stay well represented purchases - visits.
We have about the same visits by segment.
The standard deviation on young big city is high that implies that this segment is homogeneous.
Th standard deviation say us that male small city is more heterogenius.

The height of each bar represents the mean shop visits.
The vertical line, on the other hand, indicates the dispersion of the data points or how big the standard deviation is.

With that in mind, we can see that the fewer opportunities segment visits the store least often while the career focused visits it most.
That is least alike when it comes to how often they visit the shop

It seems to me that the female small city and young big city clusters are very similar in terms of their average store purchases.



In a comparative of every brand with the choice on every segment the brand 2 is the favorite on male small city and female small city segments, and brand 5 and 4 is the favorite of young what live on big city.

Looking Revenue

Segment	Revenue Brand 1	Revenue Brand 2	Revenue Brand 3	Revenue Brand 4	Revenue Brand 5	Total Revenue	Segment Proportions
young big city	1198.52	2396.60	1255.45	13163.21	19919.77	37933.55	0.314
female small city	2042.45	11702.38	711.07	2659.42	6968.36	24083.68	0.362
male small city	3064.68	7669.33	4055.00	3217.47	2732.87	20739.35	0.324

Young big city bring the mos revenue followed by female small city, and last male small city. Young big ciy is the smallest part of the proportion on the dataset, this mean that buy the most expensive soda.

Comparison female and male, female have the second revenue but they are more people (4% more than male) but the revenue is more than 4%, it's the 20%, thats mean that females buy the sodas more cheapers.

And male buy soda of all prices.

The Brand 3 have the worst revenue from all, can be interesting reduce the price of this for increase the sales.

The brand 4 have only sales from young big city, we can increase the price a little for increse the benefit, the young big city will continue buy this soda, only buy this and 5
We do all with the objetive of retain our customers and increase the revenue per sales.

Purchase Analytics and Predictive Analysis

A logistic regression is a classification method which outputs a probability between 0 and 1.

This output could be used in two different ways.
First, we use case as a probability estimate.
If we get 0.77 as an output, we would consider that there is 77% chance of purchase.
The second prominent use case is as a classifier, so if the output number is below 0.5, we usually classify it as zero.
In our case, that would translate to no purchase.
Alternatively, if the output number is above 0.5, we classify as one or purchase.
We'll still need to load our data and import the three pickled objects Scaler, PCA and K-means with PCA.
We'll preprocess our data in the same way as we did in the descriptive Statistics notebook.
Our goal is to obtain the same data frame as purchase predictors from our descriptive analysis.

We'll define the variables dependent and independent, our dependent variable will be Incidence, as we want to predict the purchase probability for our customers.
It's common that the price is strong influence by the possibility of purchase.
The independent vars will be the mean across the five prices(five brands of soda).

First, we want predict if there is going to be a purchase.
We are goin to use sag as solver, a solver is used to determine the maximum or minimum value of a cell by modifying other cells, (optimize the results).

I generate the model, and I get coefficients for price, it's negative, this means that if the price increases the purchase probability decreases.

An exploration of the prices

	Price_1	Price_2	Price_3	Price_4	Price_5
count	58693.000000	58693.000000	58693.000000	58693.000000	58693.000000
mean	1.392074	1.780999	2.006789	2.159945	2.654798
std	0.091139	0.170868	0.046867	0.089825	0.098272
min	1.100000	1.260000	1.870000	1.760000	2.110000
25%	1.340000	1.580000	1.970000	2.120000	2.630000
50%	1.390000	1.880000	2.010000	2.170000	2.670000
75%	1.470000	1.890000	2.060000	2.240000	2.700000
max	1.590000	1.900000	2.140000	2.260000	2.800000

It informs the price range, for which we will be exploring purchase probability.

The elastic on price is definite by this formula

$$\underbrace{E}_{\text{Elasticity}} = (1 - \underbrace{Y}_{\text{Purchase probability}}) * \overbrace{\beta_1}^{\text{Logistic regression coefficient of Price}} * \underbrace{P}_{\text{Price}}$$

We defined a range [0,5 – 3,5] for study the purchase probability

The elasticity going plain till 1- 1.25 going deeper on the next prices.
If the value of elasticity on absolute value is less than 1 is inelastic, otherwise elastic.

That mean that if for example we change the price to 1.0 Eur it's on the range of values |E| < 1 theorical does not affect to the sales because it's on range.(inelastic)

But if for example raise to 2 the elasticity is -3,5 could decline the purchase probability in 3,5 %. (elastic)

If our elasticity is low we can increase our prices without fear loss sales, but if our elasticity is high we should have a lower price.

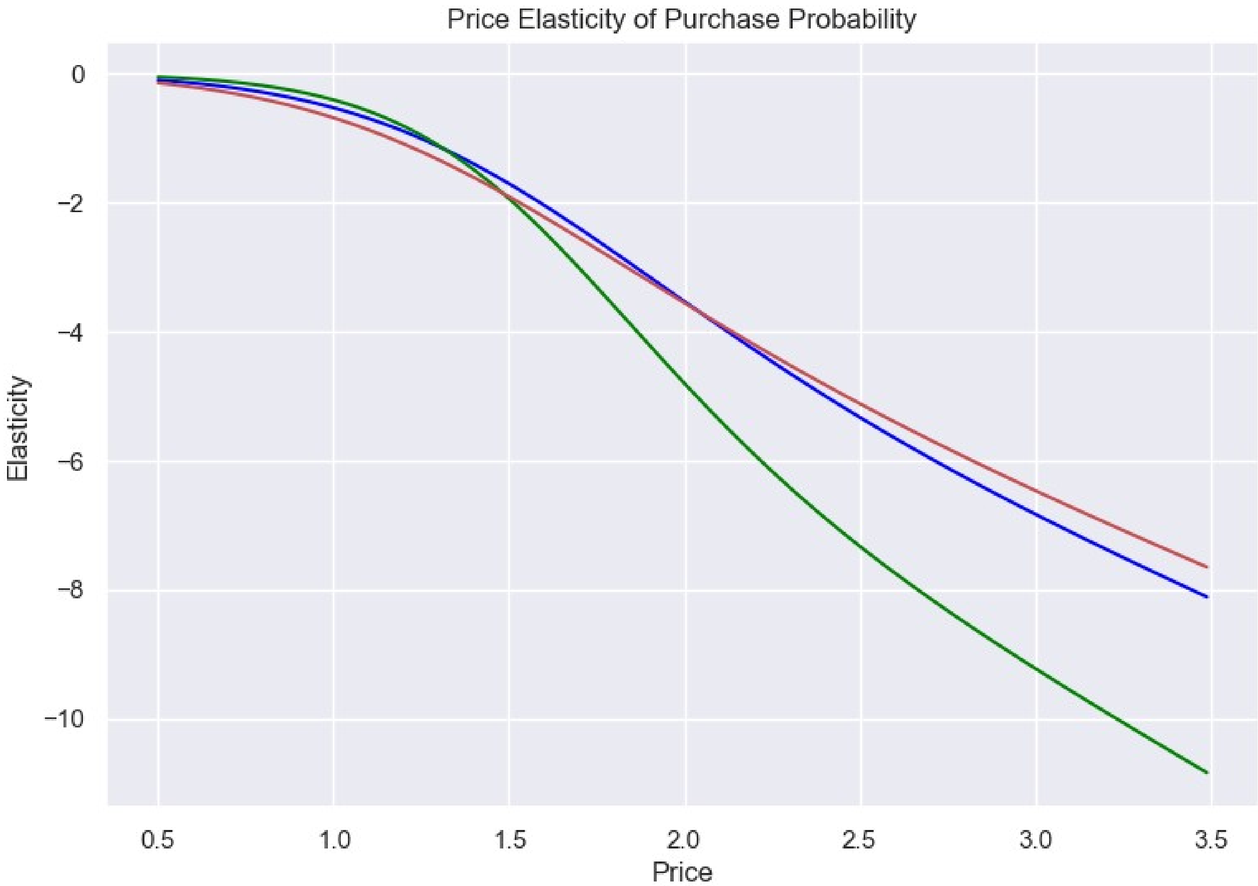
Purchase Probability by Segments

Segment 1 - female small city
Calculate the coeficient of logistic regresion.
We observe that the price coefficient for the female small city is -3.11, whereas for the average customer it was -2.35.
This is an indicator, that this segment is more price sensitive compared to the average customer.

```
array([[-3.11479769]])
```

female small city is more sensitive to raises of prices, even the probabilities falls over price of buy many of products with high prices and is more sensitive than the average of customers.

Segment 2 - male small city



Segment 2 - male small city is on average to raises of prices, even it’s better than average.

Segment 0 - young big city

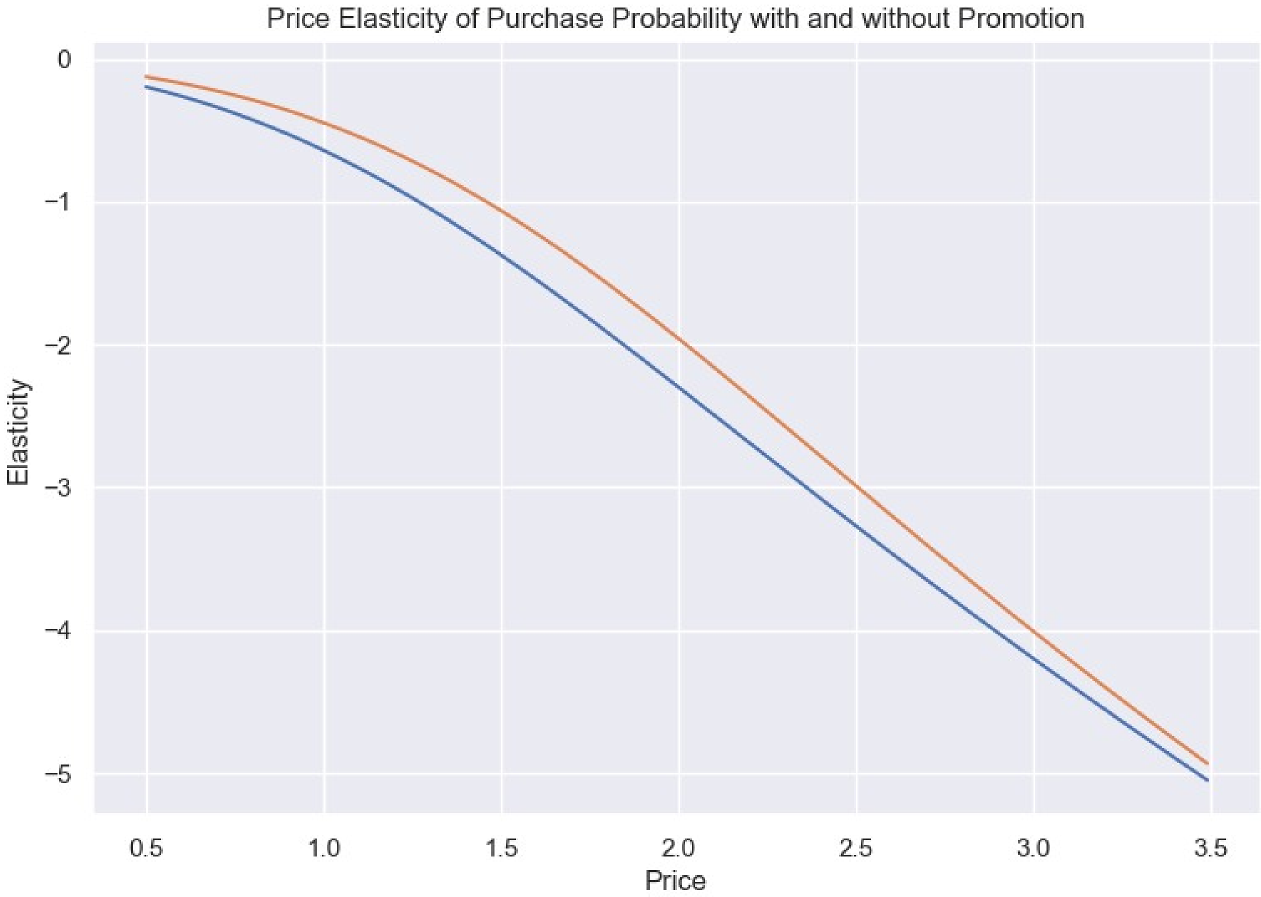
We observe that the young big city segment are the least elastic when compared to the rest, my be a segment with high acquisitive power in other hand we have female small city that is very sensitive to the rises of prices. And male small city is on average, but sometimes under the average, sometimes upper, may be a segment heterogeneous.

Purchase Probability with Promotion Feature

The promotion may be affect to the purchase intention.

We include a second promotion feature. We'd like to examine the effects of promotions on purchase probability.
We calculate the average promotion rate across the five brands. We add the mean price for the brands.

An calculate with no promotion



We observe elasticity curve with promotion and no promotion promotion is the orange and no promotion is the blue.
There are similar, sometimes put product on promotion, but vary on some cents for skyrocket sales.

Next steeps….

Apply tensorflow and deep learning to make predictions about all the areas in digital marketing and “analogic” marketing